# Weekly report

## 1 Done

### 1.1 Make a plan for the privacy-preserving map project.

### 1.2 Provide some advices for Singapore students.

### 1.3 Learn Angular from website.

### 1.4 Catch a cold.

### 1.5 Read papers:

**Locally Densest Subgraph Discovery**

(From KDD 2015)

This paper provides a parameter-free definition of an LDS with several useful properties, which can be computed in polynomial time. Furthermore, three novel optimization strategies to improve the algorithm are introduced

Definitions:

1) $\text{density}(G) = \frac{|E(G)|}{|V(G)|}$
2) ρ-compact: A graph G is ρ-compact if and only if G is connected, and removing any subset of nodes S ⊆ V(G) will result in the removal of at least ρ × |S| edges in G, where ρ is a nonnegative real number.
3) Locally Densest Subgraph: A subgraph g of G is a locally densest subgraph (LDS) of G if and only if g is a maximal density(g)-compact subgraph in G.

The algorithm can be summarized as:

**Algorithm 2** LDS(graph $G$, integer $k$)

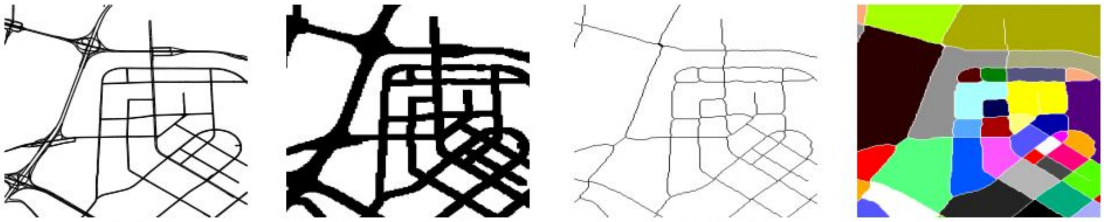
```
G' ← G;
for i = 1 to k do
    find ← false;
    while not find and G' ≠ ∅ do
        g ← any connected component of Densest(G');
        G' ← the residual graph of G' after deleting g;
        if Verify(g, G) then { find ← true; output g; }
Procedure Verify(subgraph g, graph G)
g' ← TryDensity(G, density(g) − 1/|V(G)|^2);
return g is a connected component in g';
```

A straightforward greedy approach finds the densest subgraph at a time, removes it from the graph and repeats this procedure for k times. However, such approach may not fully reflect the top-k densest regions of a graph, moreover, the subgraph returned by this approach can be partial and subsumed by a better subgraph.

**Discovering Regions of Different Functions in a City Using Human Mobility and POIs**

(From KDD 2012)

A raster-based map is a binary image (e.g., 0 stands for road segments and 1 stands for blank space). Firstly, the unnecessary details, such as the lanes of a road and the overpasses are removed by dilation operation (thicken the roads). Secondly, the skeleton of the road networks are obtained by thinning operation. Finally, a connected component labeling (CCL) is performed according to clustering grids.



(a) Original (b) Step 1: dilation (c) Step 2: thinning (d) Step 3: CCL

There are three steps to implement the territory identification, which including region aggregation, functionality intensity estimation and region annotation. In the part of functionality intensity estimation, people's mobility patterns are regarded as a clue, thus KDE model is used to evaluate based on the data about the pick-up/drop-off points of taxies.

**An Optimized Rubber-Sheet Algorithm for Continuous Area Cartograms**

(From The Professional Geographer 2013)

With the condition for topology preservation, the key optimization to the algorithm is to design and calibrate a force generation equation that can meet such a condition.

In summary, Opti-DCN keeps the algorithmic logic and main steps of the original rubber-sheet algorithm and it optimizes DCN in three critical ways. First, the points used to generate forces are changed from all vertices to points on circles and squares. Second, the force generation equation in Opti-DCN is a scaled negative exponential function. Last, a global elasticity coefficient is calculated from the Jacobian matrix of the transformation function and can largely guarantee topological integrity and optimal convergence rate. Because the differential of an exponential function is itself, calculating the new, optimized global elasticity coefficient adds little computational burden to Opti-DCN.

**A fast, free-form rubber-sheet algorithm for contiguous area cartograms**

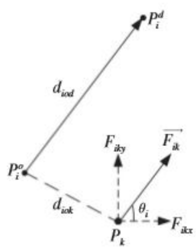
(From International Journal of Geographical Information Science 2013)

Based on the definition of a set of transforming points pairs like $P_i^o$ and $P_i^d$, another point $P_k$ is affected as:

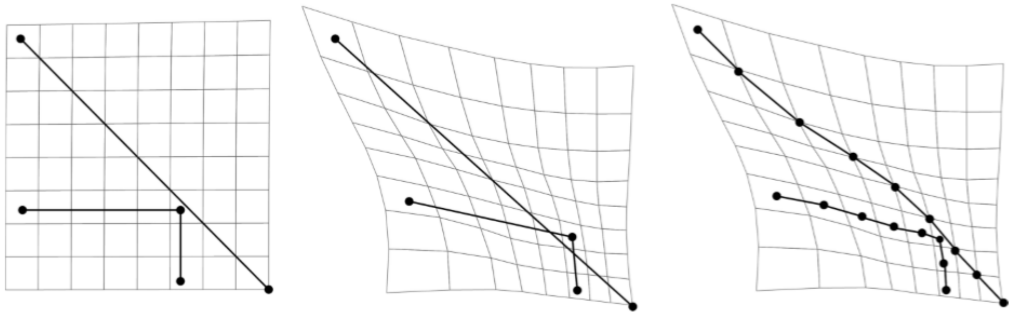
$$\overrightarrow{F_k(x_k, y_k)} = F_{k\cdot x}(x_k, y_k)\overrightarrow{u_x} + F_{k\cdot y}(x_k, y_k)\overrightarrow{u_y}$$

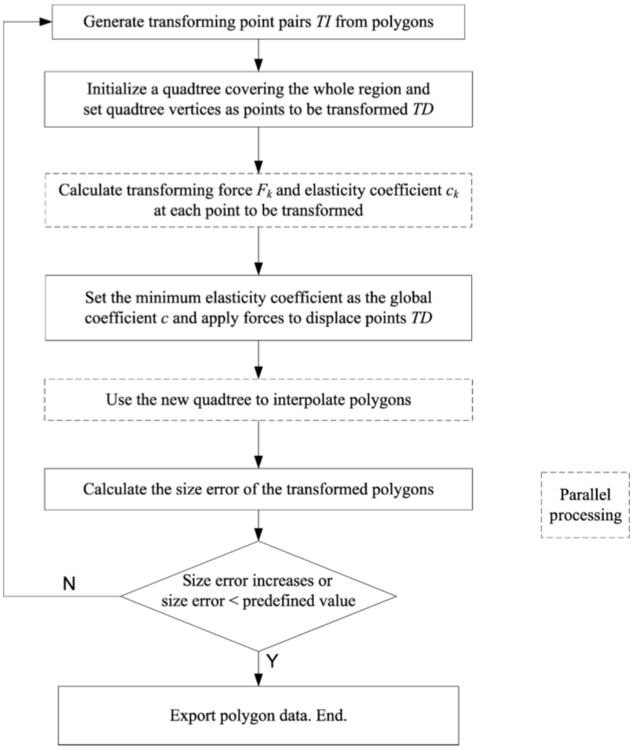
$$F_{kx}(x_k, y_k) = \sum_{i=1}^{n} F_{ikx} = \sum_{i=1}^{n} \left(x_i^d - x_i^o\right) e^{-\frac{d_{iok}}{d_{iod}}}$$

$$F_{ky}(x_k, y_k) = \sum_{i=1}^{n} F_{iky} = \sum_{i=1}^{n} \left(y_i^d - y_i^o\right) e^{-\frac{d_{iok}}{d_{iod}}}$$



To preserve the topology, the auxiliary full quadtree is applied. The reasons why they choose quadtree are that the quadtree structure can systematically eliminate the negative impacts of long arcs without significantly increasing the computational cost and quadtree itself is a spatial indexing and compression structure.






**Privacy-Preserving High-Quality Map Generation with Participatory Sensing**
(From INFOCOM 2014)

To eliminate the possible risk of privacy exposure, one natural way is to let the user report fewer locations information. However, to address the debacle between map quality and user privacy, in this work, they let each user update a subset of location points (which are randomly shuffled to remove the temporal ordering of points in the trace) to the server so as to minimize a certain measure of map generation errors.